DOCUMENT RESUME

ED 389 185 FL 023 383

AUTHOR Greis, Naguib
TITLE Bridging the Evaluation Gap in ESL.
PUB DATE 95
NOTE 12p.; Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (29th, Long Beach, CA, March 26-April 1, 1995).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *English (Second Language); *Language Tests; Peer Evaluation; Second Language Instruction; Self Evaluation (Individuals); *Student Evaluation; *Test Format; *Testing Problems; Test Wiseness

ABSTRACT

A discussion of testing in English-as-a-Second-Language (ESL) instruction focuses on the gap between ESL students' test performance on the one hand and their own and teachers' assessments of their competence on the other. First, a number of issues, drawn from the literature, are examined briefly, including the appropriateness of current testing methods, the teacher's role in evaluation, and variation in learners' test-taking strategies and skills that affect evaluation. Three ways to improve ESL student evaluation are explored: (1) clarification of evaluation criteria and use of criterion- vs. norm-referenced tests; (2) student participation in the assessment process, through self-monitoring and peer evaluation; and (3) improved use of technology, both to tailor test items to the learner's level and to provide useful feedback. Contains 36 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Naguib Greis
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Naguib Greis
Portland State University

ED 389 185

Bridging the Evaluation Gap in ESL

Introduction

An essential component in the design of the ESL curriculum, whether notional-functional, skill-based or content-based, is assessment of the learners' academic and social needs (Dubin and Olshtain 1986:25; Krahnke 1987:75; Katz 1988:178). These needs cannot be fully met without the proper evaluation of the learners. Students feel motivated when their evaluation gives them a sense of fairness and acceptablity.

However, there is often a wide gap between the students' performance on tests and their perceived competence. The gap is further widened by the difference between the teacher's evaluation and the one based on standard tests. While some ESL programs have partly recognized the learners' needs by adding a testing course, no serious attempt has been made to address the evaluation gap.

Aim and Scope

The aim of this article is first, to examine test techniques, strategies and evaluation procedures in ESL, and second, to suggest how these can be effectively integrated into the ESL program to bridge the evaluation gap. For the learners to achieve an accurate perception of their performance level, techniques such as self-monitoring and regular peer feedback are described. While the testing system is scrutinized, the thrust of the article is to help the learner function within the current system. The discussion will focus on the learners as active participants and on the teacher's role to guide them to take full responsibility for their learning and performance (Kabonen 1992:36).

Basic Terms

To clarify the discussion, the basic terms should be first explained. **Evaluation**, as used here, refers to the interpretation of student performance on tests with due attention to test characteristics and individual needs and goals. **Assessment** is measurement that is often used as the first step toward evaluation. **Tests** are to be distinguished from exercises and examinations in that they are designed to measure specific skills utilizing established formats such as multiple choice and the cloze procedure. A criterion-referenced

BEST COPY AVAILABLE

test "measures a student's performance according to a particular standard or criterion" whereas a norm-referenced test shows how the student's performance compares with that of another group used as the norm. (J.D. Brown 1993:168). The literature refers to test writers, test users and test takers.

Other distinctions to be kept in mind include achievement vs. proficiency and, from the learner's perspective, grade vs. score. The definition of **proficiency** varies from one writer to another. It ranges from "the ability to communicate accurately" along a continuum(Lowe & Stansfield 1988:13) to obtaining a specific score on a proficiency test (Jones 1981:107).

Relating Modern Methodology to Testing and Evaluation
Although testing is an integral part of the learning process, methods of testing and evaluation have not kept up with the changes in learning methods and approaches. Very often testing is treated as a separate field with little or no reference to the reactions or psychology of the learner (Buck 1994:147). The emphasis seems to be more on the statistical procedures than on the learners' benefits (James 1981:43).

Criticism is usually aimed at the structure or content of tests rather than the procedure of evaluation . For example, Gattegno, in defending the Silent Way, feels that "...we rarely focus upon our students' actual progress and instead measure their distance from a pre-established end where they 'should be' presumably because of our means and approaches" (1978: 201). He advocates what he calls "continuous feedback, the backbone of correct assessment and evaluation" (p.199).

Another learning approach advocate, Wilkins, suggests a different method of testing. He explains (1976:82) that in the Notional approach

> ...we will be seeking the answer to the question of whether the learner can express such things as concepts of time, spatial relationships, possibilities, intentions, promises, forgiveness, prohibitions, affirmations, conjectures, surprises, solicitude - indeed any of the sub-categories that are proposed for the notional syllabus.

He, however, admits that at the time of writing his book " we do not know how to establish the communicative proficiency of the learner" .

Neither Gattegno nor Wilkins deals specifically with the evaluation procedure and that is probably true of many other advocates of modern teaching approaches. At the same time, we find a great deal about the need to adopt modern learning methods which usually emphasize two major charcteristics, namely individualization and collaborative learning.

The learning process has been described as participatory, communicative, cooperative, experiential and interactive. According to Nunan (1992:4), one of the three areas of collaborative learning is progress monitoring and evaluation tasks. But as far as the actual testing and evaluation procedures are concerned, there is still a gap rather than collaboration between the learners and the teachers.

The gap is further widened by the diversity among the learners who vary in their perception of their level of competence; some lack confidence and some are overconfident. They also vary in their test-wiseness, skills, motivation, level of anxiety and general attitude toward testing. In some cases, when they see the vast difference between the tests and the class activities, they tend to look for short cuts such as studying mainly for the test or even cheating in some situations.

The Role of the Teacher

The teacher, on the other hand, in trying to encourage the students may give "generous" grades or evaluations regardless of the learners' actual performance. Grades are partly based on classroom considerations such as attendance, participation and test results (Heilenman 1990: 188).

This situation is further complicated by the difference between the teacher-made tests that are based on a specific textbook and standard tests such as TOEFL and Michigan. From the perspective of ESL students, grades roughly stand for the teacher's evaluation whereas scores are ultimately the basis for the Admissions officer's evaluation. The students are usually aware of the admission policy in many institutions to reduce evaluation to a single score such as 500 or 550 on the TOEFL (Thomas 1994:328).

In this system, where students do not understand the evaluation procedures, let alone taking part in them, they become teacher-dependent and that may seriously affect their motivation and

progress. But here the teacher is in a unique position to address the problem. There will have to be a balanced evaluation based on a proper understanding of the complex factors involving the tests and the learners. Above all, the learners must be educated about the tests, their benefits and limitations as they are guided to participate in the evaluation process.

This is a process that requires both individualized and cooperative work involving the students and the teacher. At the same time, the teacher should take into consideration the wider issues in testing and language acquisition (Swain 1993:203). While paying attention to the characteristics of tests, their validity and reliability, the teacher should also be aware of the developing nature of second language acquisition.Thus the rate of progress is usually faster in beginners than in advanced learners and at times these learners may reach a plateau (Spolsky 1990:12).

It is also important that the teacher should be sensitive to the influence of the learners' backgrounds and interactive abilities on their performance on tests (Duran 1984:45). The interaction can be structured so as to allow participation within the learners' "zone of personal development" to use Vygotsky's terms (Mohan & Smith 1992:98).

Clearly the role of the teacher is quite challenging. It involves integrating teaching and testing, exposing the learners to a variety of test formats and providing the proper interpretation of test results. Another responsibility is facilitating feedback and interaction to cultivate the learners' effective strategies while they are involved in self- and peer-evaluation.

The Learners' Strategies

With their different cultural and educational backgrounds, the learners are bound to vary in their styles, strategies and skills which affect their performance on tests. Some are recognized as 'slow' learners and some as 'fast' learners. Then there are those that tend to be analytic while others tend to be holistic and communicative-oriented (Ellis 1990:94) Furthermore, teachers often distinguish between those who rely heavily on memory and those who prefer more social strategies considered typical of specific national groups such as the Middle Easterns, Asians and Hispanics.

It has also been observed that successful learners are likely to use a variety of strategies and the advanced ones to use more task-relevant strategies (Oxford 1990:104). By identifying the students' individual strategies, the teacher can take the first step toward helping them to cultivate effective strategies. In this respect, the weaker students can benefit from the teacher's explanations and from their peers' feedback.

As far as test-taking is concerned, some students are skilled in test-wiseness (TW) which is defined as " the ability to use test-taking strategies to select the correct response in multiple-choice tests, without necessarily knowing the content or using the skill that is being tested " (Allan 1992:121). For example, in a test of TW, as opposed to 'normal' strategies, the choice may be based on eliminating obviously incorrect alternatives and looking for grammatical clues or a pattern of answers. In some cases, students show a tendency "to respond to factors other than question content" known as "response effects" (Heilenman 1990:175). According to Allan, "taking a test of TW [for which he provides an example in Appendix B pp.114-119] and receiving feedback might be enough to sensitize learners to the use of unfamiliar test-taking strategies" (Allan 1992:110).

Test-taking strategies may be considered in relation to the specific language areas such as listening comprehension, speaking, reading and writing. But there are basic strategies that ESL students should be trained to develop from the beginning. Among these are following directions, speed and careful timing. For those who always complain about lack of time, practice can help. Moreover, familiarity with the various formats of tests and types of questions in the various language areas should be useful.

Test Practice and Evaluation Criteria

Adequate practice on different test types and formats should be provided so as to include, for example, multiple-choice, true-false, matching checklists, completion and the cloze procedure. In the choice of test types, attention should be paid to the advantages of criterion-referenced tests as compared with norm-referenced tests (J.D. Brown 1993). Here the role of the teacher is crucial in explaining the value and limitations of test types to the students before interpreting the results as part of the evaluation procedure.

In the clarification of evaluation criteria, a great deal can be learnt from the manuals and guidelines for such tests as TSE (Test of Spoken English, TWE (Test of Written English) and TOEFL. According to Lowe & Stansfield (1988:3) the most comprehensive guidelines combine those of ETS, ACTFL and ILR (Interagency Language Roundtable) . Interviews can be videotaped and evaluated by students to the extent possible in the light of the appropriate TSE criteria. Similarly, students can discuss their writing (organization and ideas) based on the TWE six-point scale. It is important to be aware of the complexity of the factors involved and the need to understand the bases for judgment (Douglas 1994). To participate in proper evaluation a certain amount of training is obviously necessary.

Self- evaluation and Feedback

As indicated earlier, to bridge the evaluation gap the students must be guided to participate in the evaluation process both individually and as a group. Instead of the traditional practice where the teacher knows best or as Stevick put it "Now try to do this so I can tell you how you did " (Quoted by Haughton and Dickinson 1988:234), the idea is to become self-directed and responsible for one's learning and performance on tests.

There are several reasons for self-evaluation. As already mentioned, it is a step toward an accurate perception of one's ability. Furthermore, it is not only a "necessary part of self-direction", but also "one way of alleviating the assessment burden on the teacher" (Dickinson 1987:136).

But the question is sometimes raised regarding the accuracy and reliability of self-evaluation. While some studies show a significant correlation between self-assessment and objective assessment, others indicate little or no correlation. Others still believe that both self-evaluation and external evaluation complement each other (de Bot 1992:138).

The process of self-evaluation entails adopting techniques for self-monitoring in every language area (Oscarson 1989). These include self-reports, diaries, questionnaires, checklists and charts. In reading, for example, charts may be kept by the student to show progress in speed, comprehension and vocabulary. To evaluate their writing, the students may discuss in small groups criteria for organization and cohesion similar to those used in TWE as already pointed out. Along

the same lines, suggestions based on TSE may be used for evaluation. While general standard tests cannot be reviewed, samples may be carefully studied.

Feedback from students can significantly enhance the process of evaluation. Students in small groups may discuss the performance of individual members in the light of criteria and relevant information provided by the teacher. As part of the feedback process, students record their reactions to tests indicating how they respond, the level of difficulty and what they like or dislike about the questions.

Such feedback may be provided through questionnaires and short questions along a 5-point scale. Comments can indicate the students' view in terms of fairness and acceptability (A. Brown 1993:278-9). Besides, the mere attention to student comments is likely to be motivating and, as Madsen *et al.* (1991: 66) point out, "reflects an interest in the total process and not simply in the intellect or skill mastery".

This participation in the evaluation process has additional advantages. Apart from enabling the students to achieve a more accurate perception of their performance level and enhancing their motivation, it can reduce their anxiety level. Excessive test anxiety can be debilitating especially as the degree of anxiety usually increases with the degree of evaluation perceived (Daly 1991:9).

The Role of Technology

Modern technology can help facilitate the process of testing and evaluation. Two promising areas are the latent trait theory or item response theory (IRT) and computerized adaptive testing (CAT). IRT assumes according to Tung "an accurately calibrated set of items that assess a single dimension of the examinee's ability" (Tung 1986:27). The difficulty and discrimination power of an item vary according to the level of the examinee. IRT has been known since 1986 but it was only after its application in CAT that it became valuable (Stansfield 1986:5). Thus it was possible to tailor test items to the examinee's level. Although CAT has its limitations in dealing with complex communicative activities, it may be used in diagnostic testing of such discrete areas as vocabulary and sentence structure.

Furthermore, computers can be a valuable tool in providing immediate feedback regarding progress evaluation with accuracy and speed (Carroll & Hall1985:135).. They can provide regular

reports that the student may discuss with other students and with the teacher (J.D. Brown 1993:180-181). As a result, through computer-assisted testing, which lends itself to individualized work, self-evaluations with actual performances may be compared and the information shared by learners and teachers. Such information has implications for the "learners' perceptions of their abilities"(Alderson 1990:26).

Conclusion

Bridging the evaluation gap in ESL involves individualized as well as cooperative work. Three major areas are considered, namely guided test practice, self-evaluation and feedback. In their practice on various types of tests, the students are encouraged to learn from each other and cultivate effective strategies.

As a counselor or facilitator, the teacher structures the class activities so as to integrate testing and evaluation into the program. The learners are provided with the opportunity to be active participants in the evaluation process guided by discussion, explanations and peer feedback. Not only will the process reduce their test anxiety and enhance their attitude and motivation, but it will also help them achieve a more accurate perception of their abilities.

Further research is needed to assess the impact of changes in the evaluation procedure on bridging the gap. Self-evaluation may be compared with the teacher's evaluation and examined in light of student feedback.

References

Alderson, J.C. (1990). Learner-centered testing through computers: International issues in individual assessment. In J.M.L. Jong & D.K. Stevenson (Eds.). *Individualizing the assessment of language abilities.*, (pp. 20-27). Clevendon, Philadelphia: Multilingual Matters LTD.

Allan, A. (1992).Development and validation of a scale to measure test-wiseness in EFL/ESL reading test tasks. *Language Testing,* 9,101-122.

Baker, D. (1989).*Language testing*: A critical survey and practical guide. London and New York: Edward Arnold.

Brown, A. (1993).The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing,* 10, 277-303.

Brown, J.D. (1993). A comprehensive criterion-referenced testing project. In D. Douglas & C. Chapelle (Eds.) *A new decade of language testing research* : Selected papers from the 1990 language testing research colloquium, (pp,163-184). Washington,DC: TESOL.

Buck, G. (1994).The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing,* 11, 145-170.

Carroll, B. and P.J. Hall (1985). *Make your own language test:* A practical guide to writing your own language tests. Oxford & New York: Pergamon Institute of English.

Daly, J. 1991. Understanding communication apprehension: An introduction for language educators. In E. Horwitz & D.J. Young (Eds.) *Language anxiety*: From theory and research to classroom implications, 3-13. Englewood Cliffs, NJ: Prentice Hall.

de Bot, K. (1992). Self-assessment of minority language proficiency. In L. Verhoeven & J.H.A.L. de Jong (Eds.) *The construct of language proficiency:* Applications of psychological models to language assessment, (pp.137-146). Amsterdam/Philadelphia: John Benjamins Publishing Co.

de Jong, J.H.A.L. & D.K. Stevenson (1990). *Idividualizing the assessment of language abilities.* Clevendon, Philadelphia: Multilingual Matters LTD.

Dickinson, L. (1987). *Self-instruction in language learning*. New York: Cambridge University Press.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing*, 11, 125-144.

Dubin, F. & E. Olshtain 1986. *Course design : Developing programs and materials for language learning*. Cambridge: Cambridge University Press.

Duran, R.P. (1984). Some implications of communicative competence research for integrative proficiency testing. In C. Rivera (Ed.) *Communicative competence approaches to language proficiency assessment: Reserarch and application.*, (pp.44-58). Clevendon, Avon, England : Multilingual Matters LTD.

Ellis, R. 1990. Individual learning styles in classroom second language development In J.H.A.L. de Jong & D.K. Stevenson (Eds.) *Individualizing the assessment of language abilities*, (pp. 83-96). Clevendon, Philadelphia: Multilingual Matters LTD.

Gattegno, C. Evaluating students' progress. In C.H. Blatchford, & J. schachter (Eds.) *On TESOL* '78, (197-202).

Haughton, G. and Dickinson, L. (1988). Collaborative assessment by masters' candidates in a tutor based system. *Language Testing*, 5, 233-246.

Heilenman, K. L.(1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7, 174-201.

James, C. J.(1981). Language testing as a key to language learning. In J.E. Redden (Ed.) *Proceedings of the Southern Illinois language testing conference*, (pp.43-56). Carbondale, Illinois: Southern Illinois University.

Jones, L. R. (1981). Assessing Second language proficiency: Where are we and where are we going? In J.E. Redden (Ed.) *Proceedings of the Southern Illinois language testing conference*, (pp.103-115). Carbondale, Illinois: Southern Illinois University.

Katz, A. (1988). The academic context. In P.Lowe, Jr. & C.W. Stansfield (Eds.) *Second language proficiency assessment: Current issues.* Englewood Cliffs, N J: Prentice Hall Regents.

Kabonen, V. (1992).Experiential language learning: Second language learning as cooperative learner education. In D. Nunan (Ed.) *Collaborative language learning and teaching*, (pp. 14-39), New York: Cambridge University Press.

Krahnke, K. 1987. *Approaches to syllabus design for foreign language teaching*. Englewood Cliffs, NJ: Prentice-Hall, INC.

Lowe, P.,Jr., & C.W. Stansfield. (Eds.) (1988).*Second language proficiency assessment: Current issues*. Englewood Cliffs, N J: Prentice Hall Regents.

Madsen, H.S., B.I. Brown & R.L. Jones.(1991). Evaluating student attitudes toward second language tests. In E.K. Horwitz & D.J. Young (Eds.) *Language Anxiety:* From theory and research to classroom implications, (pp. 65-86). Englewood Cliffs, N.J. : Prentice Hall.

Mohan, B., & S. Smith. (1992).Context and cooperation in academic tasks. In D. Nunan (Ed.) *Collaborative language learning and teaching*, (pp.81-99). Cambridge: Cambridge University Press.

Nunan, D. (1992) (Ed.). *Collaborative language learning and teaching*. New York: Cambridge University Press.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6, 1-13.

Oxford, R. (1990). Styles, strategies, and aptitude: Connections for language learning. In T.S. Parry & C.W. Stansfield (Eds.) *Language aptitude reconsidered*, (pp. 67-125). Englewood Cliffs, N J : Prentice Hall Regents.

Price, M. 1991. The subjective experience of foreign language anxiety: Interviews with highly anxious students. In E. Horwitz & D.J. Young (Eds.) *Language anxiety*: From theory and research to classroom implications, (pp.3-13). Englewood Cliffs, NJ: Prentice Hall.

Spolsky, B. (1990). Social aspects of individual assessment. In J.H.A.L. de Jong & D.K. Stevenson (Eds.) *Individualizing the assessment of language abilities*, (pp.3-15). Clevendon, Philadelphia: Multilingual Matters LTD.

Stansfield, C. W. (1986). *Technology and language testing*. Washington D.C. : TESOL.

Swain, M. (1993). Second language testing and second language acquisition: Is there a conflict with psychometrics? In *Language Testing* ,10, 193-207.

Thomas, M. (1994). Assessment of L2 proficiency in second language acquisition research. *Language Testing*, 11, 307-336.

Tung, P. (1986). Computerized adaptive testing: Implications for language test developers. In C.W. Stansfield (Ed.) *Technology and language testing*, (pp. 12-28). Washington D C: TESOL.

Wilkins, D.A. (1976). *Notional syllabuses*. Oxford: Oxford University Press.